SELF-MADE-ATC

TECHNICAL DEEP DIVE:
ADAPTING AUTOMATIC SPEECH RECOGNITION TO NEW AIRSPACES & APPLICATIONS

Matthias Kleinert

Niclas Feier

SELF-MADE-ATC

DLR

Automatic Speech Recognition and Understanding (ASRU)



ASRU adapted for Vienna Approach 24 person months of effort



austrian four seven two tango
good morning wien radar
identified follow balad three
november transition



Performance:
Word Error Rate = 3%
Command Recognition = 92%
Callsign Recognition = 99%

Automatic Speech Recognition and Understanding (ASRU)



Use ASRU without modifications for Zurich En-Route





edelweiss four four bravo
climb flight level three three
zero increase mach point seven
five



~~Performance:~~
~~Word Error Rate = 3%~~
~~Command Recognition = 92%~~
~~Callsign Recognition = 99%~~

Benefits of SELF-MADE-ATC



How to improve the performance?

~~"Just" Contact DLR and ask them to fix it → **With another X months of effort**~~
~~"Just" hand over 50 hours of speech from Zurich en-route traffic → **Do not forget to stay GDPR-conform**~~
~~"Just" Manually transcribe 10 hours of speech data → **Easily takes 100 hours of expert working time**~~
~~"Just" manually annotate 10 hours of the transcriptions → **Again takes 100 hours of expert working time**~~

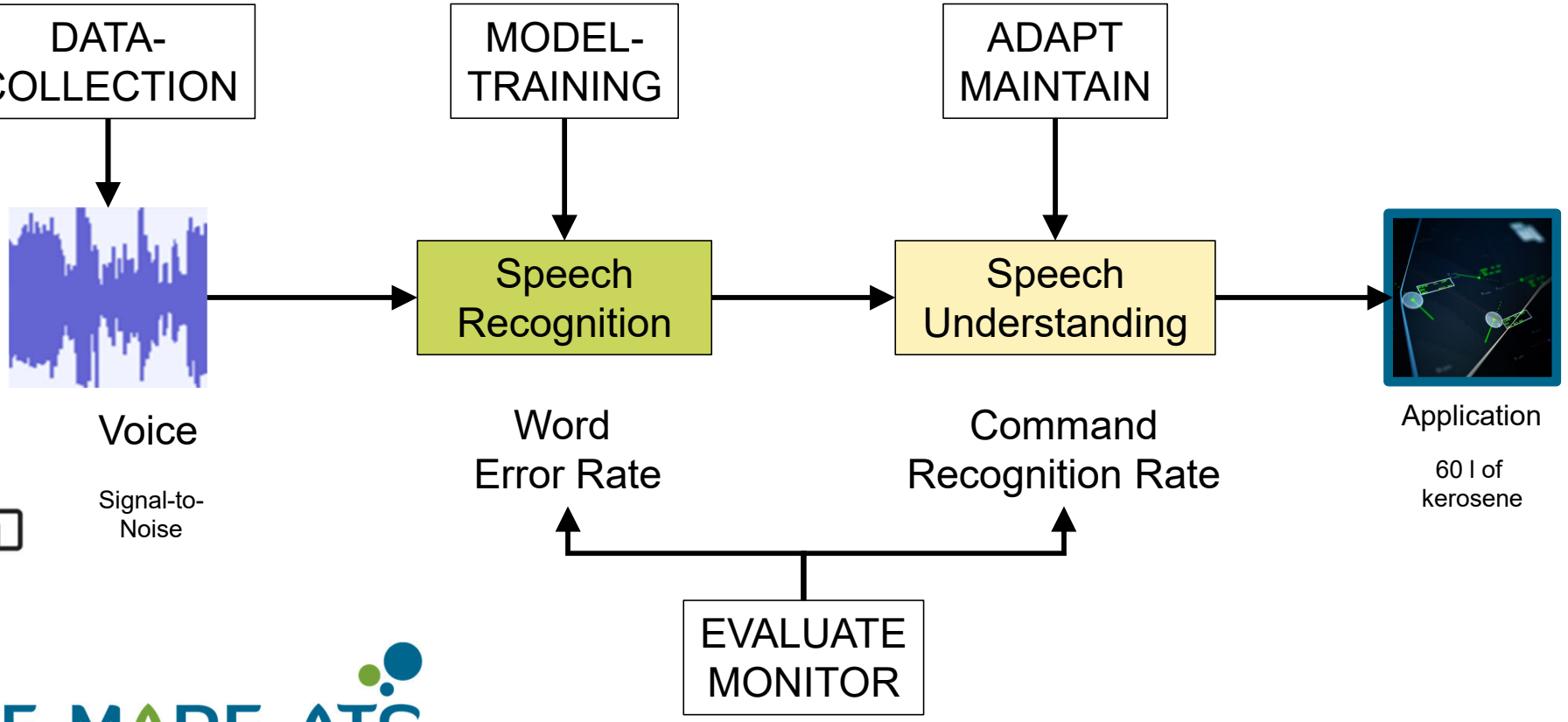
↓

SELF-MADE-ATC

↓

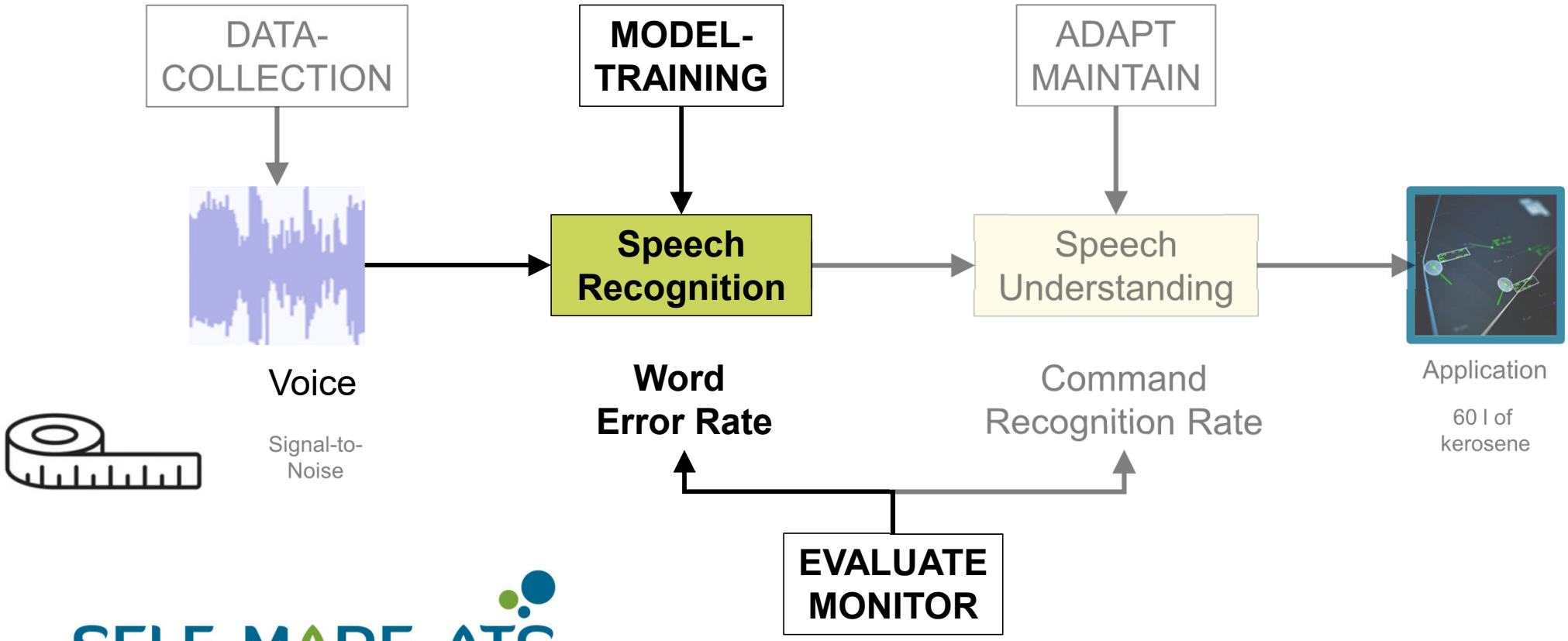
"No" need of DLR support (just as backup) → **Just need one month of adaptation effort**
"No" data sharing with third parties – data stays within the ANSP/airport facilities → **No GDPR trouble**
"No" manual transcription (only for evaluation) → **Only minimum amount of expert working time**
"No" manual annotation (only for evaluation) → **Only minimum amount of expert working time**

Building Blocks for Adapting ASRU



SELF-MADE-ATC

Building Blocks for Adapting ASRU



SELF-MADE-ATC

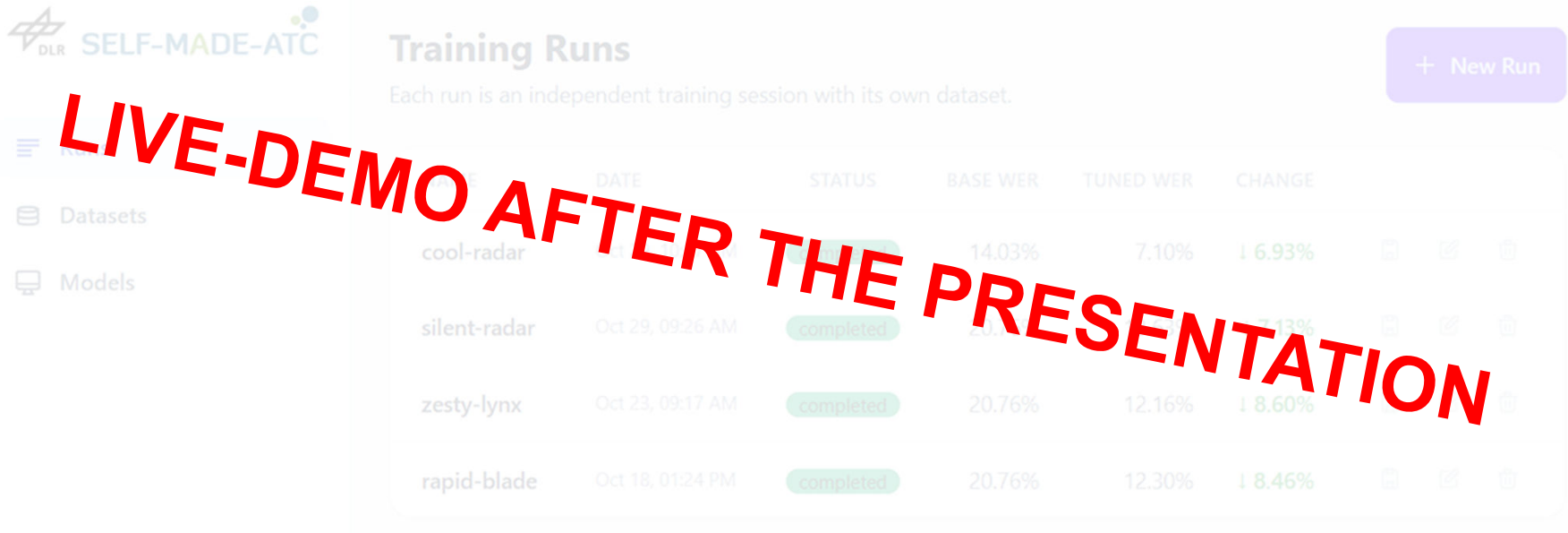
Model Training – Automation Pipeline

- Automated pipeline that adapts speech recognition models to your desired airspace
- The only thing you need to take care of → Provide your own data
  - **Training** → Voice Transmissions + Context (optional) – **<u>No manual transcription effort</u>**
  - **Test** → Voice Transmissions + Transcriptions – **<u>Minimal manual effort for evaluation</u>**



Model Training – Automation Pipeline

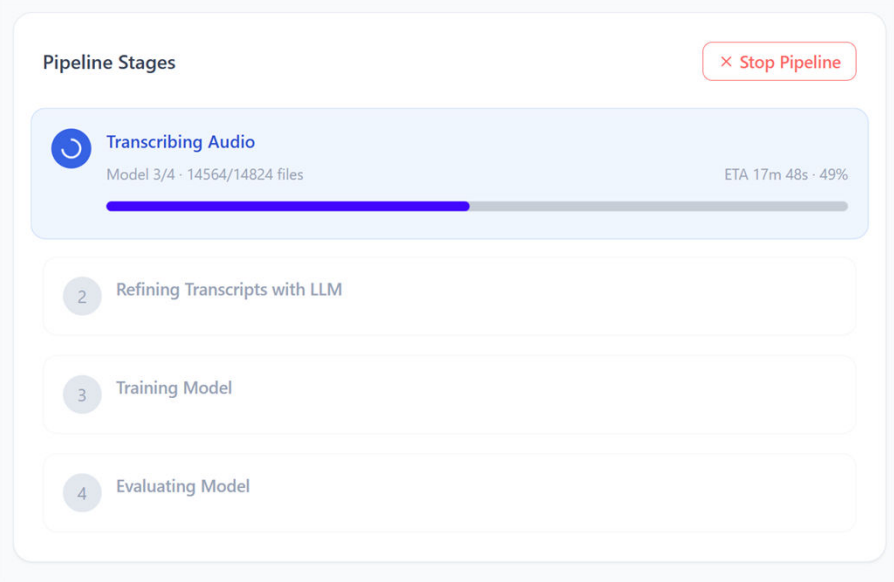
- Automated pipeline that adapts speech recognition to your airspace → Easy to use GUI
- The only thing you need to take care of → Provide your own data
  - **Training** → Voice Transmissions + Context (optional) – **No manual transcription effort**
  - **Test** → Voice Transmissions + Transcriptions – **Minimal manual effort for evaluation**



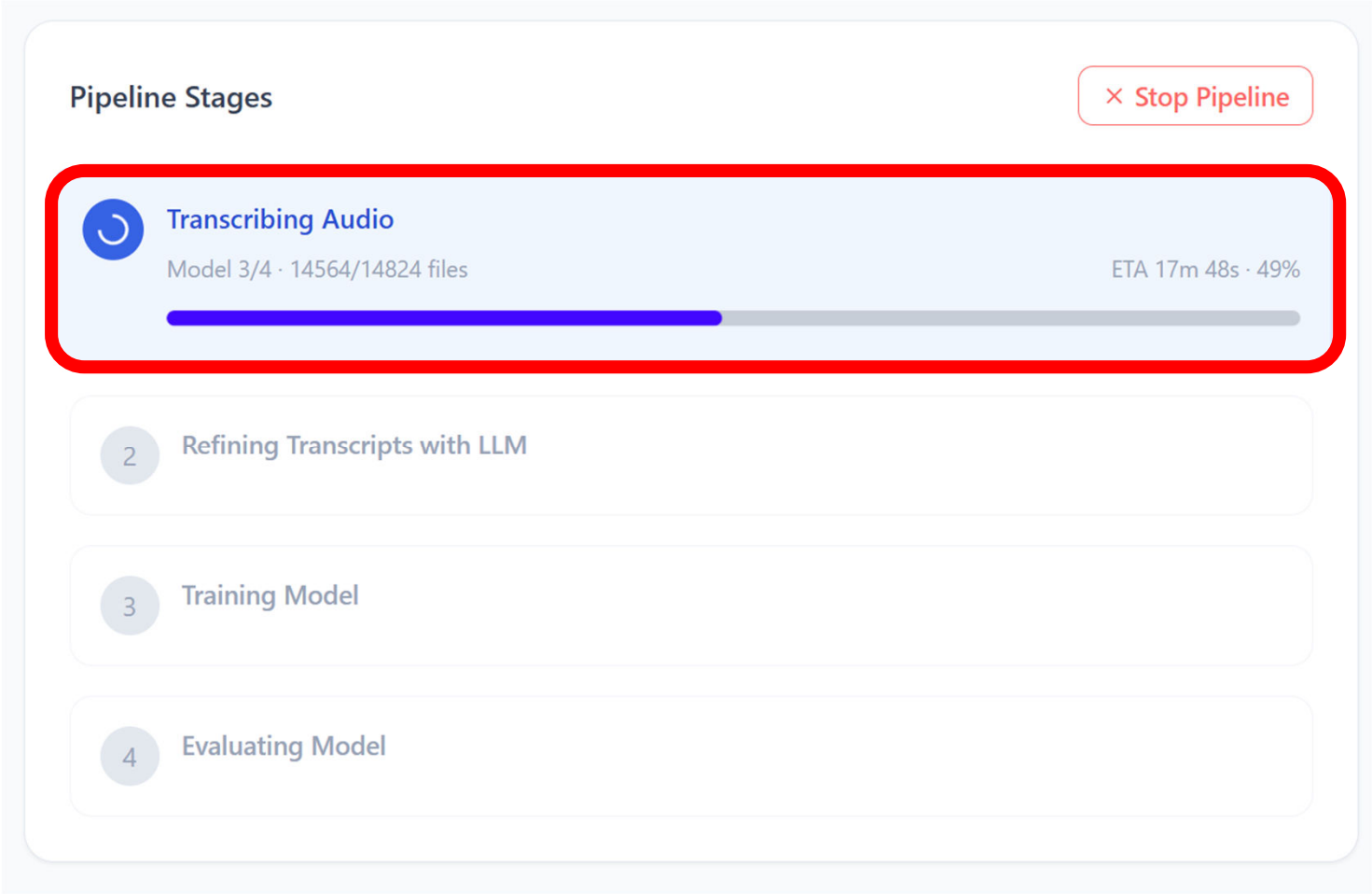
LIVE-DEMO AFTER THE PRESENTATION

Overview

- For the rest of the presentation we assume the following:
  - You uploaded some data to the SELF-MADE-ATC pipeline
  - You defined a SELF-MADE-ATC pipeline run that uses the uploaded data
  - You executed the pipeline run



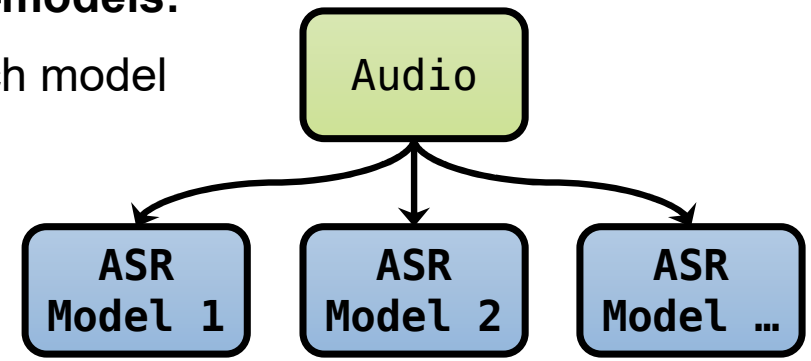
Overview


Pipeline Stages
× Stop Pipeline
Transcribing Audio
Model 3/4 · 14564/14824 files
ETA 17m 48s · 49%
2
Refining Transcripts with LLM
3
Training Model
4
Evaluating Model

Transcribing Audio

Our approach relies on a combination of different ASR-models:

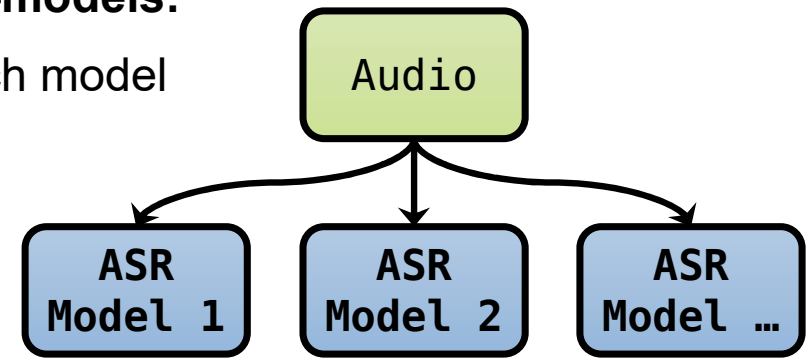
- Voice transmissions are transcribed independently by each model
- Intentionally different training data and architectures
- Different rather than correlated error patterns



Transcribing Audio

Our approach relies on a combination of different ASR-models:

- Voice transmissions are transcribed independently by each model
- Intentionally different training data and architectures
- Different rather than correlated error patterns



Example – Transcripts

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

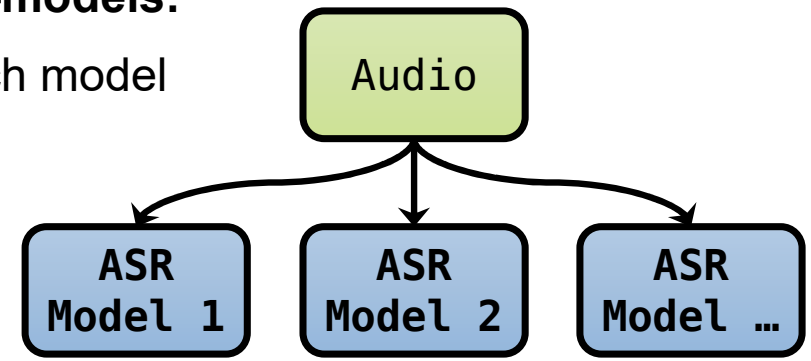
M-3 → radar five four mike bravo descending level three four zero level by korep

→ Substantial variations in vocabulary, casing, numerical representations and punctuation

Transcribing Audio

Our approach relies on a combination of different ASR-models:

- Voice transmissions are transcribed independently by each model
- Intentionally different training data and architectures
- Different rather than correlated error patterns



Example – Transcripts

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

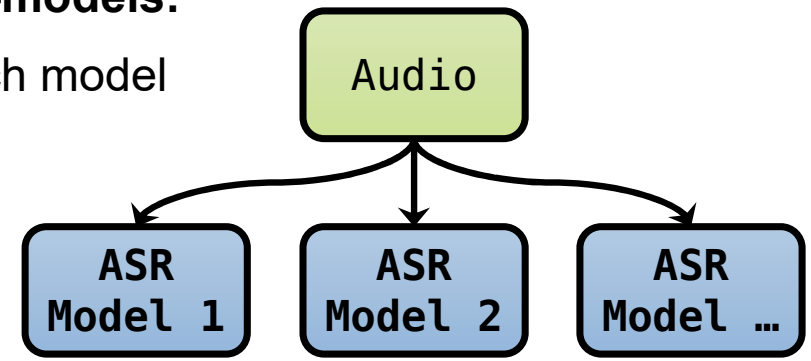
M-3 → radar five four mike bravo descending level three four zero level by korep

→ Substantial variations in vocabulary, casing, numerical representations and punctuation

Transcribing Audio

Our approach relies on a combination of different ASR-models:

- Voice transmissions are transcribed independently by each model
- Intentionally different training data and architectures
- Different rather than correlated error patterns



Example – Transcripts

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

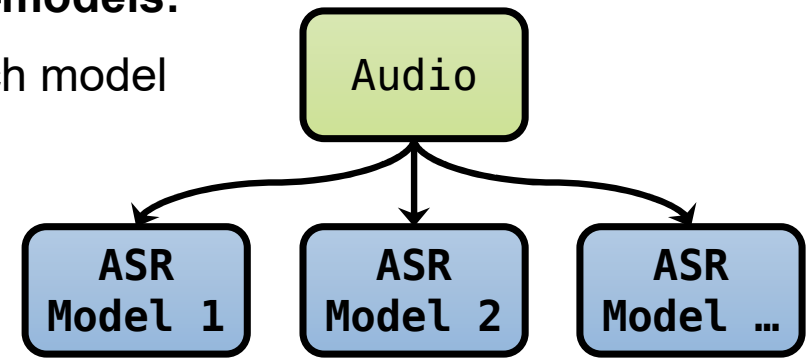
M-3 → radar five four mike bravo descending level three four zero level by korep

→ Substantial variations in vocabulary, casing, numerical representations and punctuation

Transcribing Audio

Our approach relies on a combination of different ASR-models:

- Voice transmissions are transcribed independently by each model
- Intentionally different training data and architectures
- Different rather than correlated error patterns



Example – Transcripts

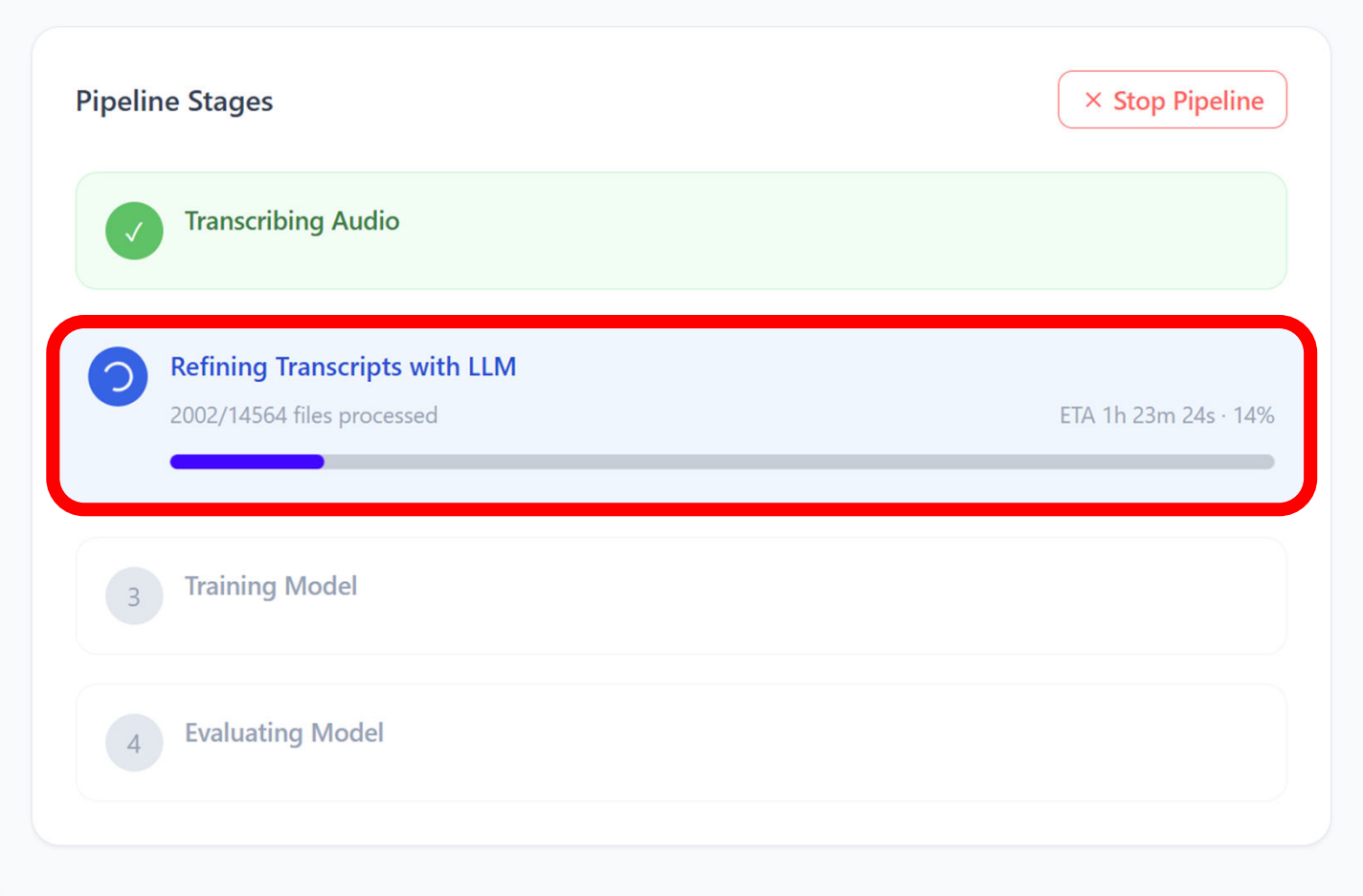
M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

→ Substantial variations in vocabulary, casing, numerical representations and punctuation

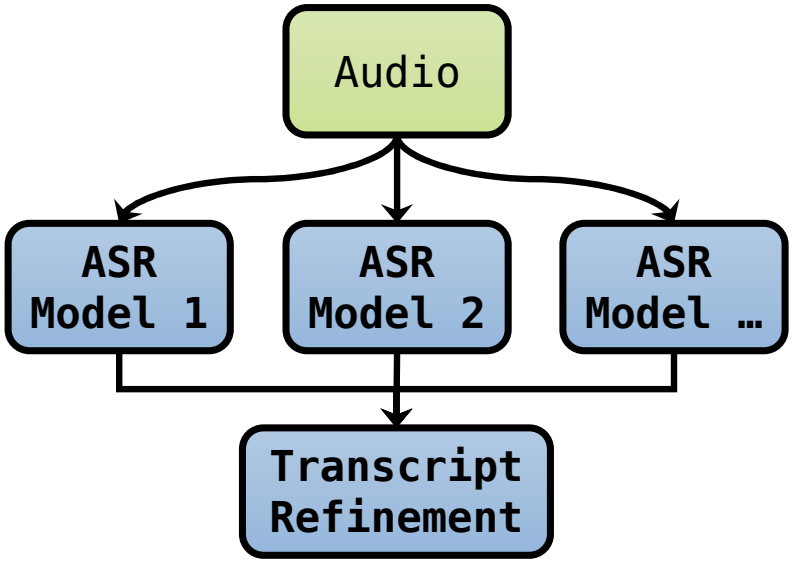
Overview


Pipeline Stages
× Stop Pipeline
Transcribing Audio
Refining Transcripts with LLM
2002/14564 files processed
ETA 1h 23m 24s · 14%
3
Training Model
4
Evaluating Model

Transcript Refinement with LLM

ASR-Model output dropped in LLM-based fusion process:

- The LLM's role → Act as a human subject matter expert
- Its task → Create a single ground truth transcript
- Task prompt defines:
  - Fusion Rules – How to fuse the input transcripts
  - Context Handling – How to handle context information
  - Few-Shot Examples – Concrete examples of fusions
  - Output Contract – How the output should be presented



Transcript Refinement with LLM – Fusion

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREEK

M-3 → radar five four mike bravo descending level three four zero level by korep

- LLM now has to build a ground truth transcript based on these ASR-Ensemble Input:

Transcript Refinement with LLM – Fusion

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

- LLM now has to build a ground truth transcript based on these ASR-Ensemble Input:

ryanair five mike bravo descending level three four zero level by korep

Transcript Refinement with LLM – Fusion

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

- LLM now has to build a ground truth transcript based on these ASR-Ensemble Input:
  - Some words are picked because obviously all models agree

ryanair five mike bravo descending level three four zero level by korep

Transcript Refinement with LLM – Fusion

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

- LLM now has to build a ground truth transcript based on these ASR-Ensemble Input:
  - Some words are picked because obviously all models agree
  - Others are chosen because of a majority decision

ryanair five mike bravo descending level three four zero level by korep

Transcript Refinement with LLM – Fusion

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

- LLM now has to build a ground truth transcript based on these ASR-Ensemble Input:
  - Some words are picked because obviously all models agree
  - Others are chosen because of a majority decision
  - And some are picked purely because they seem to be the most reasonable

ryanair five mike bravo descending level three four zero level by korep

Transcript Refinement with LLM – Fusion

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR LIMA FIVE MIKE BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

- LLM now has to build a ground truth transcript based on these ASR-Ensemble Input:
  - Some words are picked because obviously all models agree
  - Others are chosen because of a majority decision
  - And some are picked purely because they seem to be the most reasonable

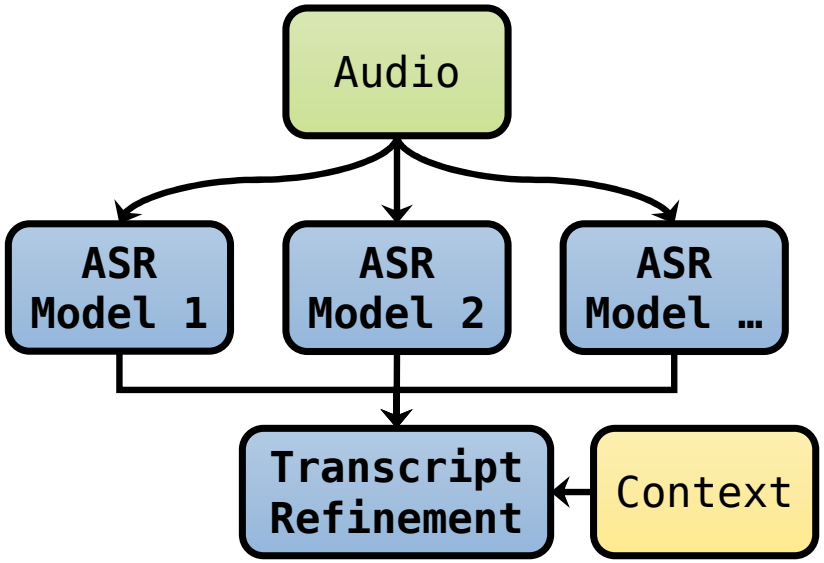
ryanair five mike bravo descending level three four zero level by korep

→ The resulting transcript is better than each individual one

Transcript Refinement with LLM – Context is Valuable

To enhance the LLM output the pipeline injects operational context

- Context helps to detect and resolve misrecognitions and ambiguities
- Two distinct classes
  - Static airspace context: Waypoints
  - Dynamic surveillance context: Callsigns



Transcript Refinement with LLM – Context is Valuable

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR FIVE MIKE LIMA BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep



ryanair five mike bravo descending level three four zero level by korep

Transcript Refinement with LLM – Context is Valuable

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR FIVE MIKE LIMA BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

↓

ryanair five mike bravo descending level three four zero level by korep

↓

Callsigns: RYR5MLB = ryan air five mike lima bravo, …
Waypoints: KORET, …

Transcript Refinement with LLM – Context is Valuable

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR FIVE MIKE LIMA BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

↓

ryanair five mike bravo descending level three four zero level by korep

↓

Callsigns: RYR5MLB = ryan air five mike lima bravo, …
Waypoints: KORET, …

↓

ryan air five mike lima bravo descending level three four zero level by koret

Transcript Refinement with LLM – Context is Valuable

M-1 → Right, hello, under 5 for my Bravo, descending level 340, level 540.

M-2 → RYANAIR FIVE MIKE LIMA BRAVO DESCENDING LEVEL THREE FOUR ZERO LEVEL BY KOREK

M-3 → radar five four mike bravo descending level three four zero level by korep

↓

ryanair five mike bravo descending level three four zero level by korep

↓

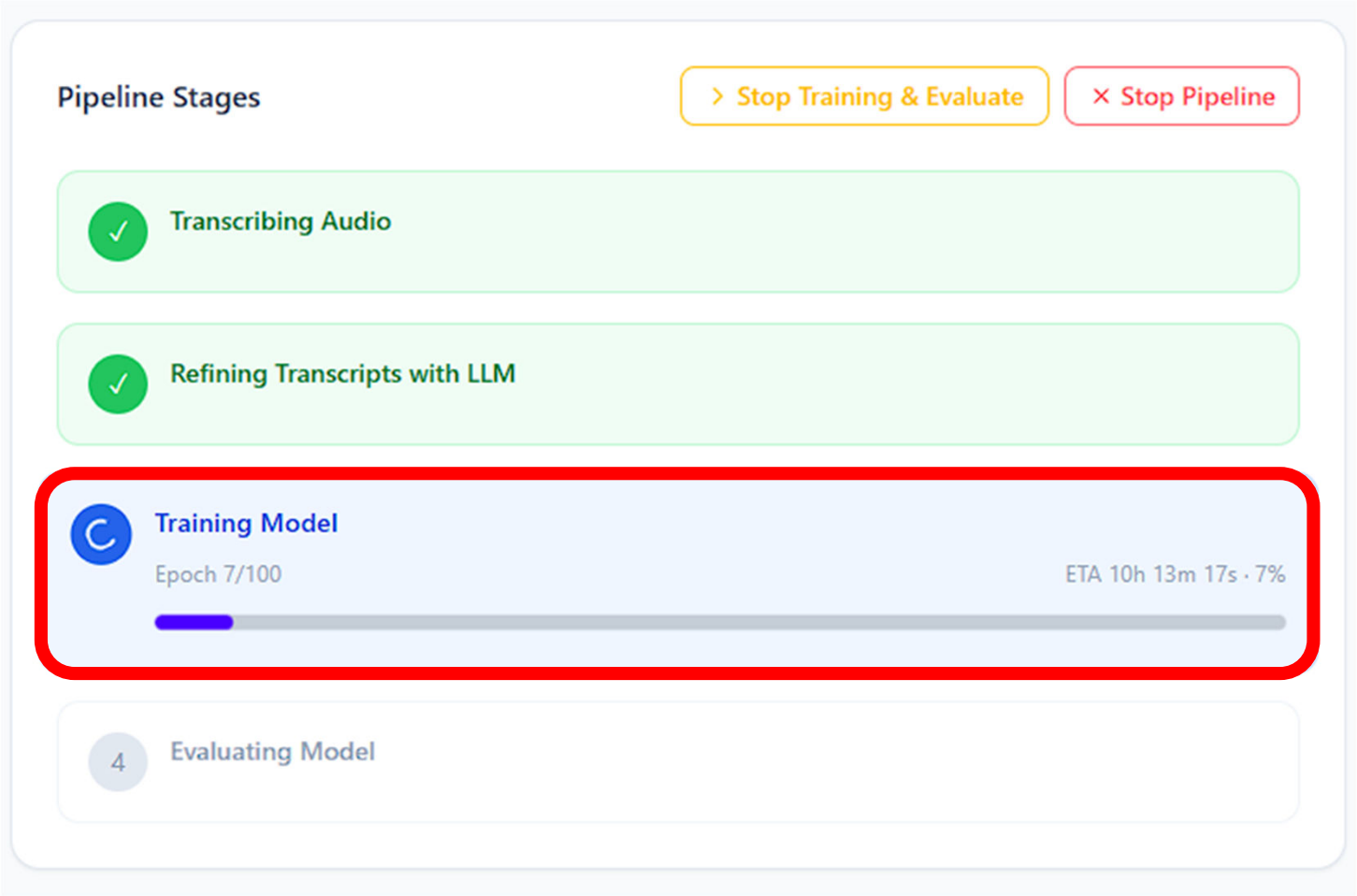
Callsigns: RYR5MLB = ryan air five mike lima bravo, …
Waypoints: KORET, …

↓

ryan air five mike lima bravo descending level three four zero level by koret

→ We just produced a high-quality transcript for ATC audio with no human labour

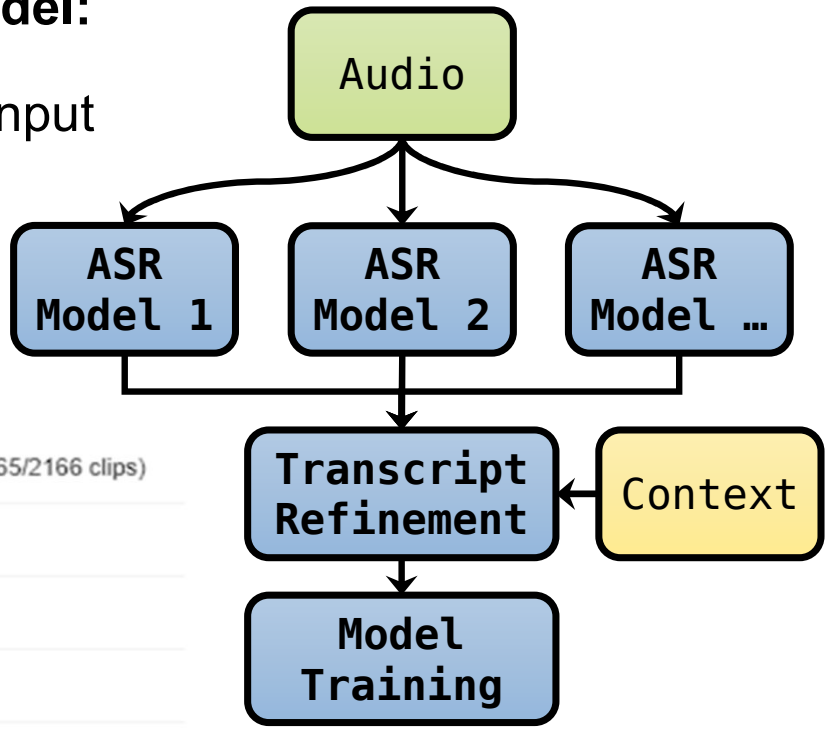
Overview

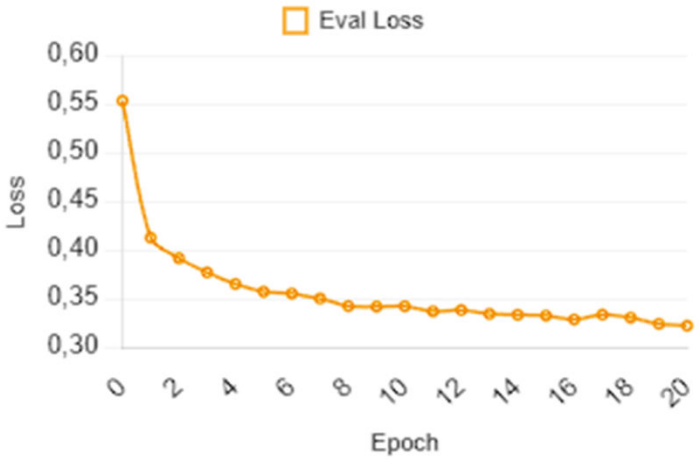

Pipeline Stages
Stop Training & Evaluate
Stop Pipeline
Transcribing Audio
Refining Transcripts with LLM
Training Model
Epoch 7/100
ETA 10h 13m 17s · 7%
4
Evaluating Model

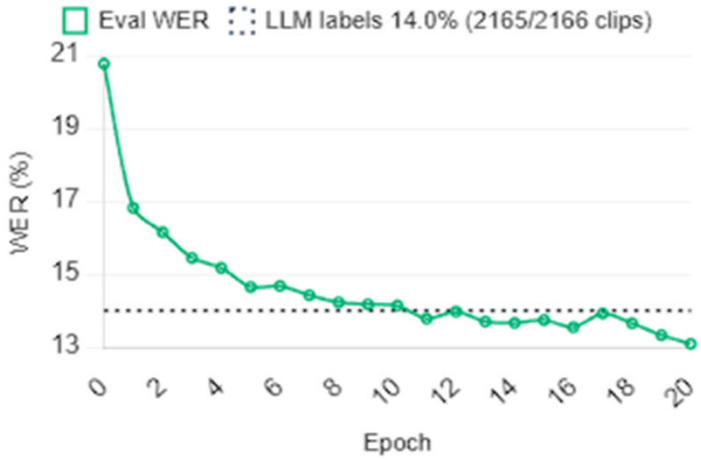
Model Training

Pipeline launches automatic adaptation of ASR model:

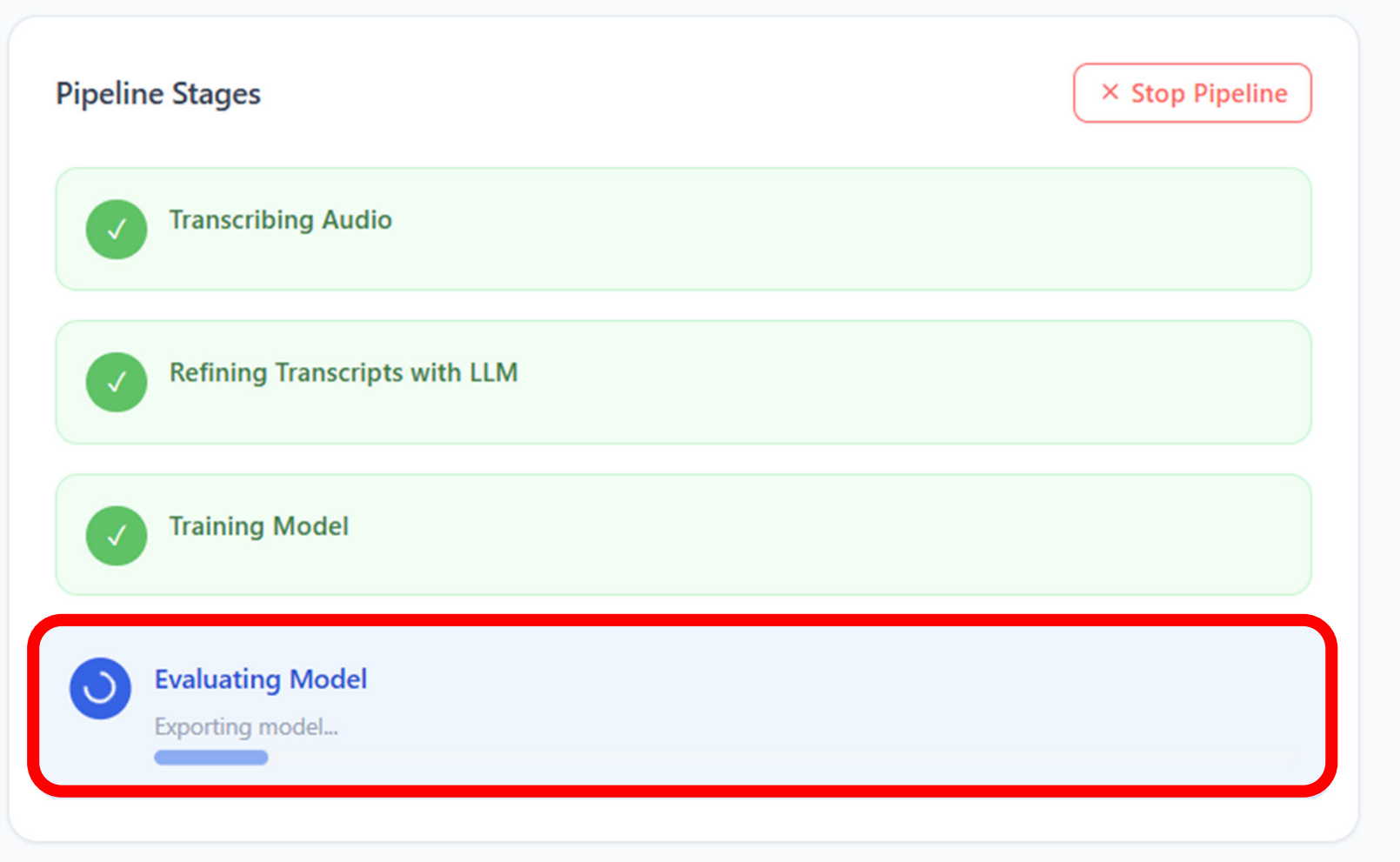
- Training uses ATC audio + generated transcripts as input
- Adapts the operational ASR model
- Provide Continous overview of the progress







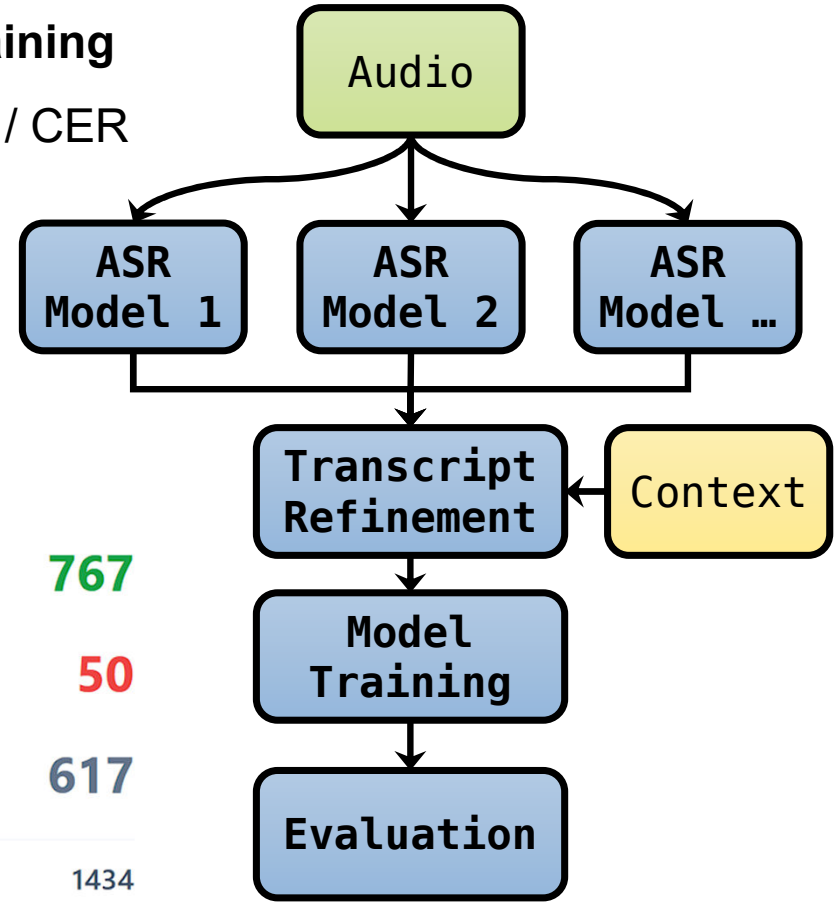
Overview

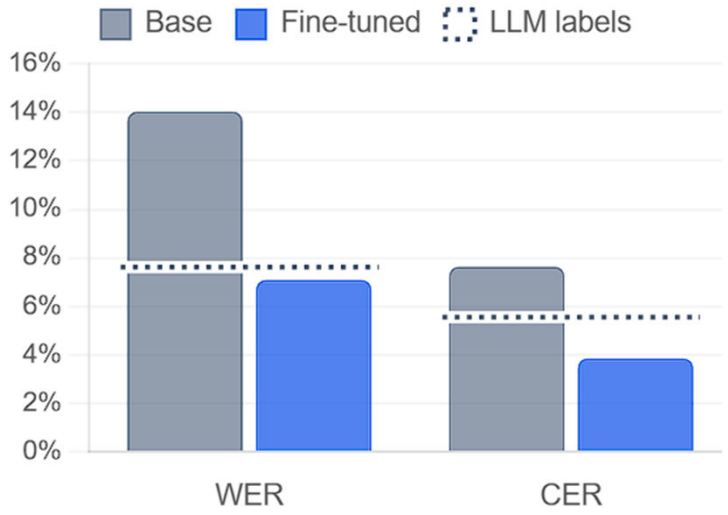

Pipeline Stages
× Stop Pipeline
Transcribing Audio
Refining Transcripts with LLM
Training Model
Evaluating Model
Exporting model...

Evaluation

Overview of the achieved performance after the model training

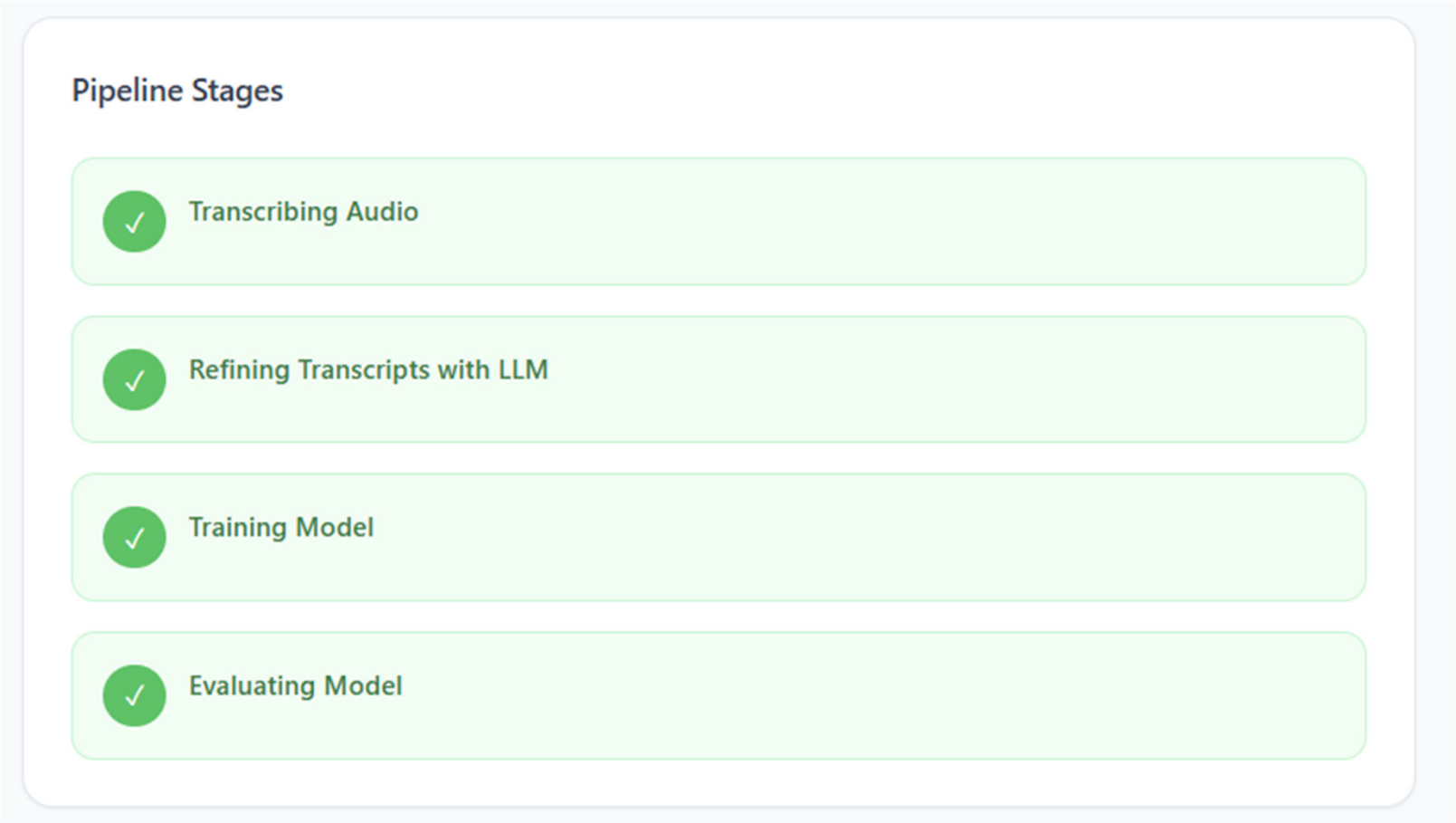
- Dedicated information about improvements regarding WER / CER
- Information about improvements per voice transmission
- Detailed anlysis on individual files







Overview


Pipeline Stages
Transcribing Audio
Refining Transcripts with LLM
Training Model
Evaluating Model

Summary

- Automation → Pipeline processes data, executes training and evaluates results automatically
- No human transcription effort for training data → Makes it scalable across large data sets
- Context-aware → Injecting contextual data helps to resolve misrecognitions and ambiguities
- Deployable → Adapted ASR model can be downloaded and deployed after training
- On-Premise Execution → No cloud connection everything runs on a single GPU machine



END

Quelle: Software-Entwickler Web-Entwickler - Kostenloses Foto auf Pixabay

Impressum

Topic:	SELF-MADE-ATC Technical Deep Dive: Adapting Automatic SPEECH RECOGNITION to New Airspaces & Applications
Date:	2026-11-05
Author:	Matthias Kleinert, Niclas Feier
Institute:	Institute of Flight Guidance (FL)
Credits:	All figures and photos „DLR (CC BY-NC-ND 3.0)“, if not otherwise specified